AFRL-AFOSR-VA-TR-2016-0097

Compressive Information Extraction A Dynamical Systems Approach

Mario Sznaier
NORTHEASTERN UNIVERSITY

02/09/2016
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
23-02-2016

2. REPORT TYPE
Final Performance

3. DATES COVERED *(From - To)*
01-06-2012 to 31-05-2015

4. TITLE AND SUBTITLE
Compressive Information Extraction A Dynamical Systems Approach

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-12-1-0271

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Mario Sznaier

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NORTHEASTERN UNIVERSITY
360 HUNTINGTON AVE
BOSTON, MA 02115 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The goal of this research was to develop a comprehensive, computationally tractable framework for addressing a broad class of problems that entail extracting information very sparsely encoded in high volume data streams. At its core was a unified vision, centered on the use of dynamical models as information encapsulators, and blending elements from dynamical systems theory, semi-algebraic geometry, sparse signal recovery and convex optimization. It included both theory development in an emerging new field compressive information extraction and an investigation of implementation issues.
Relevance to the USAF mission: As emphasized in the Technology Horizons report, flexible, provably correct autonomy is a key enabler for maintaining the superiority and expanding the capabilities of the USAF in the next two decades. Autonomous systems endowed with analysis and decision capabilities can collect data, assess intention, and if necessary, take action, while at the same time substantially reducing the required manpower and cost, vis-a-vis existing unmanned vehicles.

15. SUBJECT TERMS
compressive information extraction, hybrid system identification

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Mario Sznaier
Unclassified	Unclassified	Unclassified	UU		

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

					19b. TELEPHONE NUMBER *(Include area code)* 617-373-5364

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

DISTRIBUTION A: Distribution approved for public release

Compressive Information Extraction: A Dynamical Systems Approach

Mario Sznaier
Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
Phone (617) 373–5364, Fax (617) 373–8970
email: msznaier@ece.neu.edu

Octavia I. Camps
Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
Phone (617) 373–4663, Fax (617) 373–8970
email: camps@ece.neu.edu

January 24, 2016

Final Report, Grant FA9550-12-1-0271
Dynamics and Control Program
Air Force Office of Scientific Research
Program Officer: Dr. Frederick A. Leve, Ph.D., AFOSR/RT

Project Summary

Scope: The goal of this research was to develop a comprehensive, computationally tractable framework for addressing a broad class of problems that entail extracting information very sparsely encoded in high volume data streams. At its core was a unified vision, centered on the use of dynamical models as information encapsulators, and blending elements from dynamical systems theory, semi-algebraic geometry, sparse signal recovery and convex optimization. It included both theory development in an emerging new field –compressive information extraction– and an investigation of implementation issues.

Relevance to the USAF mission: As emphasized in the *Technology Horizons* report, flexible, provably correct autonomy is a key enabler for maintaining the superiority and expanding the capabilities of the USAF in the next two decades. Autonomous systems endowed with analysis and decision capabilities can collect data, assess intention, and if necessary, take action, while at the same time substantially reducing the required manpower and cost, vis-a-vis existing unmanned vehicles. Arguably, a major road-block to realizing this vision stems from the curse of dimensionality. Simply put, existing techniques are ill-equipped to deal with the overwhelming volume of data that needs to be analyzed in real time. This is precisely the challenge addressed by this research: development of a computationally tractable framework for robustly extracting and processing actionable information sparsely encoded in very large data sets. The long term vision was to lay the foundations for designing systems endowed with provably correct flexible autonomy, capable of making decisions in-situ, with minimal human intervention.

Contributions to Basic Science: This research effort took the first steps towards developing a new framework –compressive information extraction– that allows for robustly extracting and processing information sparsely encoded in very large data sets. At its core is a new approach that exploits an hitherto unexplored connection between information extraction and nonlinear identification. It advanced the state of the art in control theory by developing a tractable framework for robust identification/model (invalidation) of switched systems, a key component of a comprehensive control framework for hybrid systems. In addition, this research generalized to a dynamic setting the existing *compressive sensing* framework, thus substantially extending its domain of application. Further, it unveiled deep connections between the problems addressed and those arising in other branches of engineering and applied mathematics. Examples include the connection between nonlinear dimensionality reduction methods and manifold discovery (both hallmarks of machine learning) and nonlinear identification, and between rank minimization and dynamic data interpolation.

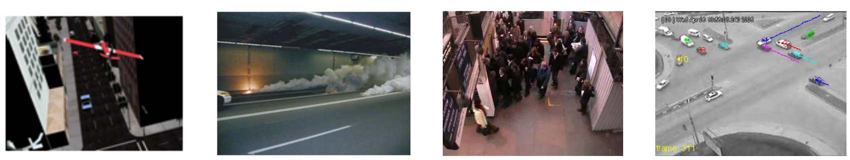
Benefits to the General Public: In addition to directly supporting the USAF mission, the results obtained in this research effort have the potential to significantly benefit society. Systems endowed with activity analysis capabilities can prevent crime, allow elderly people to continue living independently, give early warning of serious medical conditions, for instance by detecting minute gait alterations preceding a stroke, inspect aging civil infrastructures, and monitor and even coordinate responses to environmental threats to minimize their effect. Initial steps have been taken to transition the technology developed under this grant to TSA in order to enhance security at US airports. A prototype system has been installed at the Cleveland airport, where it successfully detected security breaches unnoticed by humans.

1 Motivation

The goal of this research was to develop a comprehensive, computationally tractable framework for addressing a broad class of problems that entail extracting information very sparsely encoded in high volume data streams. It was based on a unified vision, centered on the use of dynamical models as information encapsulators, that emphasized robustness and computational complexity issues. It included both theory development in an emerging new field –compressive information extraction [7]– and an investigation of implementation issues.

1.1 Transformative Impact and Relevance to the USAF Mission

As emphasized in the *Technology Horizons* report, flexible, provably correct autonomy is a key enabler for maintaining the superiority and expanding the capabilities of the USAF in the next two decades: Autonomous systems endowed with analysis and decision capabilities can collect data, assess intention, and if necessary, take action, while at the same time substantially reducing the required manpower and cost, vis-a-vis existing unmanned vehicles. Arguably, a major road-block to realizing this vision stems from the curse of dimensionality, illustrated in Figure 1. Simply put, existing techniques are ill-equipped for analyzing the "data avalanche" generated by the sensors, within the constraints imposed by the need for robust, real time operation in dynamic, partially stochastic scenarios. This was precisely the issue addressed in this project, by exploiting recent advances in robust dynamical systems, sparse signal recovery, semi-algebraic geometry and optimization. The long term vision that motivated this project was to lay the foundations for designing systems endowed with provably correct flexible autonomy, capable of making decisions in-situ, without human intervention, while passing on to the next decision level only mission–relevant situational abstractions.



(a) (b) (c) (d)

Figure 1: Examples of actionable information sparsely encoded in very large data streams. (a) Target tracking in an urban canyon; (b) and (c) sample frames showing contextually abnormal events: onset of a tunnel fire and a person entering through an exit; (d) Tracking multiple targets. In all cases decisions must be taken based on events discernible only in a small fraction (less than $\mathcal{O}(10^{-6})$) of a very large data record: the video sequences in (a)-(d) add up to megabytes, yet the useful information (a change of behavior of a single target), is contained in just a few frames.

The main idea that drove this research was to recast the problem of sparse information extraction into a hybrid systems identification/model (in)validation form. Briefly, in this approach, the observed data is treated as the output of an underlying switched dynamical system, typically represented by a difference inclusion, with jumps indicating the occurrence of events. The key observation is the fact that higher degrees of spatio-temporal correlations in the data lead to lower complexity joint models, allowing for reformulating the problem of information extraction into a *dynamic* sparsification form, which in turn can be reduced to a convex semidefinite optimization problem.

A conceptual diagram illustrating these ideas is shown in Figure 2. Notably, merely postulating the existence of a dynamically sparse underlying model led to efficient, scalable algorithms for information extraction. For instance, in this context, data can be segmented by detecting changes in suitable model invariants (such as complexity), a process that can be reduced to minimizing the rank of a matrix directly constructed from the data. Similarly, interpolating missing data and determining whether two data streams correspond to time traces of the phenomenon (for instance activity) reduces to a tractable semi-definite

Sparse Signal Recovery:

Strong prior

Signal has a sparse representation: $f(t) = \sum_i c_i \psi(t)$ with only a few $c_i \neq 0$.

Signal recovery

Sparsify the coefficients:

$$\min \|[c_1, \ldots, c_n]\|_o$$

subject to $f(t) = y(t)$.

Relax to Linear Programming

$$\min \|[c_1, \ldots, c_n]\|_1$$

subject to $f(t) = y(t)$.

Sparse Information Recovery:

Strong prior

Actionable information is generated by a low complexity dynamical system

Information recovery

Sparsify the dynamics: $\mathbf{min}_y \mathbf{rank} \mathbf{H}(\mathbf{y})$ with

$$H(y) = \begin{bmatrix} y(1) & y(2) & \ldots & y(n) \\ y(2) & y(3) & \ldots & y(m+1) \\ \vdots & \vdots & \ddots & \vdots \\ y(m) & y(m+1) & \ldots & y_{m+n} \end{bmatrix}$$

Relax to Semidefinite Programming

$$\mathbf{minTrace}\mathbf{X}(\mathbf{y}) \text{ subject to } \mathbf{L}(\mathbf{y}) \succeq 0$$

Figure 2: Sparse dynamical information recovery versus sparse signal recovery.

optimization, even in cases where the data has no time overlap.

2 Description of the Research Performed and Summary of the Results

In this section we give a brief summary of the research performed under this grant and our findings. Details can be found in the papers listed in the publications section, which can be downloaded from the Robust Systems Lab website: http://robustystems.coe.neu.edu.

2.1 Basic Science.

In principle, embedding information extraction problems in the conceptual world of systems identification makes available a rich, extremely powerful resource base, leading to computationally tractable, robust solutions. However, successful application of the ideas outlined above hinged upon the development of computationally tractable solutions to the following problems, open at the time that the project was started:

2.1.1 Robust identification of hybrid systems: As outlined before, the main idea that drove this research was to treat the observed data as the output of an underlying switched dynamical system, with events indicated by changes in invariants associated with each subsystem. In the initial phase of this research, we assumed that the data record was generated by a piecewise affine model of the form[1]

$$f\left(\mathbf{p}_{\sigma(t)}, \left\{\mathbf{x}(k), \boldsymbol{\eta}_f(k)\right\}_{k=t-i}^{t+j}\right) = 0 \tag{1}$$

where f is an affine function, the parameter vector $\mathbf{p}(t)$ takes values from a finite set indexed by piecewise

[1]Note that this can be assumed without loss of generality, since piece wise affine models are universal approximators.

constant function $\sigma(t)$ and where $\boldsymbol{\eta}_f(t)$ represents bounded noise. In this context, the information is encapsulated in the parameter vector $\mathbf{p}$. For instance, events are indicated by changes in $\mathbf{p}(t)$ (an identification problem). Similarly, two time series can be considered to have been generated by the same process if they can be explained by the same $\mathbf{p}$ (e.g. a model (in)validation problem). While both, identification and model (in)validation of switched affine systems are known to be NP-hard problems, as part of this research we have developed tractable relaxations, with optimality certificates, for two practically relevant cases:

Identification with minimum number of switches: This scenario arises for instance in fault detection, where the goal is to minimize the number of false alarms, and in segmentation problems in image processing and computer vision, where it is often desirable to maximize the size of regions (roughly equivalent to minimizing the number of boundaries). Formally, this problem can be stated as: Given input/output data $\{u_t, y_t\}_{t_0}^T$ over the interval $[t_0; T]$, and a priori information consisting of a convex set membership noise description $\mathcal{N}$ and bounds $n_u \geq n_c$ and $n_y \geq n_a$ on the order of the regressors, find a switched affine model of the form:

$$y_t = \sum_{i=1}^{n_a} a_i(\sigma_t) y_{(t-i)} + \sum_{i=1}^{n_c} c_i(\sigma_t) u_{(t-i)} + \eta_t \tag{2}$$

where u, y and η denote the input, output and noise, respectively, that explains the experimental data with the minimum number of switches. The main result of this portion of the research [2] showed that, by defining the sequence of first order differences $\boldsymbol{\delta}_t \doteq \mathbf{p}_t - \mathbf{p}_{(t+1)}$, identification with minimum number of switches can be reduced to the following sparsification problem:

$$\begin{array}{c} \min_{\mathbf{p}_t} \|\mathbf{p}_t - \mathbf{p}_{(t+1)}\|_0 \\ \text{subject to } y_t - \mathbf{r}_t^T \mathbf{p}_t \in \mathcal{N} \ \forall t \end{array} \tag{3}$$

Notably, as we showed in [2], when the noise is characterized in terms of its ℓ_∞ norm, that is $\mathcal{N} = \{\eta \colon \|\eta\|_\infty \leq \epsilon\}$, then an exact solution can be found by solving a sequence of Linear Programming problems. This is one of the very few sparsification problems where exact recovery is guaranteed, without the need for additional conditions on the data, such as decoherence.

Identification with bounded number of subsystems. In this case, the problem can be formally stated as: Given input/output data over the interval $[t_0; T]$, a bound on the ℓ_∞ norm of the noise (i.e. $\|\eta\|_\infty \leq \epsilon$) find a switched ARX model of the form (2), with no more than s subsystems, that interpolates the experimental data. Although in principle this problem is NP-hard, in the noise free case (i.e. $\eta_t = 0 \ \forall t$), it can be reduced to finding the null space of a suitable constructed matrix, followed by polynomial differentiation. The starting point to accomplish this is to rewrite (2) as

$$\mathbf{b}(\sigma_t)^T \mathbf{r}_t = 0 \tag{4}$$

where $\mathbf{r}_t = \begin{bmatrix} -y_t, y_{t-1}, \ldots, y_{t-n_a}, u_{t-1}, \ldots, u_{t-n_c} \end{bmatrix}^T$ and $\mathbf{b}(\sigma_t) = \begin{bmatrix} 1, a_1(\sigma_t), \ldots, a_{n_a}(\sigma_t), c_1(\sigma_t), \ldots \end{bmatrix}^T$, denote the regressor and (unknown) coefficients vectors at time t, respectively. The idea behind the Generalized Principal Component Analysis (GPCA) method is to decouple the identification of model parameters from the identification of the switching sequence by noting that (4) holds for some σ_t if and only if

$$p_s(\mathbf{r}) = \Pi_{i=1}^s (\mathbf{b}_i^T \mathbf{r}_t) = \mathbf{c}_s^T \nu_s(\mathbf{r}_t) = 0 \tag{5}$$

holds for all t independent of which of the s submodels is active at time t, where $\mathbf{b}_i \in R^{n_a+n_c+1}$ and $\nu_s(.)$, denote the parameter vector corresponding to the i^{th} submodel and the Veronese map of degree s, respectively. Collecting all data into a matrix form leads to:

$$\mathbf{V}_s \mathbf{c}_s \doteq \begin{bmatrix} \nu_s(\mathbf{r}_{t_o})^T \\ \vdots \\ \nu_s(\mathbf{r}_T)^T \end{bmatrix} \tag{6}$$

Hence, one can solve for a vector $\mathbf{c}_s$ in the null space of $\mathbf{V}_s$ to find the coefficients of the multivariate polynomial $p_s(\mathbf{r})$. Unfortunately, this approach breaks down in the presence of noise, since (5) no longer holds. Rather, we have the following (noisy) equivalent

$$p_s(\tilde{\mathbf{r}}_t) \doteq \prod_{i=1}^{s}(\mathbf{b}_i^T\tilde{\mathbf{r}}_t) = \mathbf{c}_s^T\nu_s(\tilde{\mathbf{r}}_t) = 0 \tag{7}$$

where $\tilde{\mathbf{r}}_t = [-y_t + \eta_t, y_{t-1}, \ldots, u_{t-1}, \ldots, u_{t-n_c}]^T$, and its associated "noisy" data matrix $\mathbf{V}_s(\mathbf{r}, \boldsymbol{\eta}) \doteq \mathbf{V}_s(\tilde{\mathbf{r}})$. The main difficulty here is that finding the coefficients of the polynomial $p(\tilde{\mathbf{r}})$ requires now finding both an admissible noise sequence $\boldsymbol{\eta}^o$ and a vector $\mathbf{c}^o$ such that

$$\mathbf{V}_s(\tilde{\mathbf{r}}^o)\mathbf{c}^o = 0 \tag{8}$$

Since $\mathbf{V}_s(\tilde{\mathbf{r}})$ is now a matrix polynomial function of the unknown noise sequence η_t, this is a computationally very challenging problem. Nevertheless, as we showed in [32], the use of polynomial optimization allows for transforming (8) to a rank minimization problem of the form

$$\begin{array}{r}\tilde{\mathbf{V}}_s(\mathbf{r}_t, \mathbf{m}^{(t)})\mathbf{c}^o = 0 \\ \text{subject to } \mathbf{M}(\mathbf{m}) \succeq 0 \text{ and } \mathbf{L}(\mathbf{m}) \succeq 0\end{array} \tag{9}$$

where all the matrices involved are affine in the optimization variables $\mathbf{m}$. At this point, a tractable convex relaxation can be obtained by using the nuclear norm as a surrogate for rank, leading to a convex semi-definite program that can be solved using widely available tools (see [24, 32] for details).

2.1.2 Extensions to missing data. In most practical scenarios only partial data is available, due for instance to occlusion or limited sensing/transmitting capabilities. In these situations, it is of interest to estimate the missing data, for instance in order to perform data association (e.g. stitch tracklets), or to uncover correlations mediated by the missing elements. We have shown that this interpolation can be reduced to a rank minimization problem, which in turn (due to its Hankel structure) can be efficiently solved using convex relaxations. These theoretical results enabled the development of a new class of algorithms, based upon polynomial optimization, capable of handling both time and frequency domain constraints, as well as constraints on the order of the interpolant [5].

2.1.3 Robust Identification of Hammerstein/Wiener Systems. In the context of information extraction, this problem arises naturally as a way of handling the extreme high volume of data involved. Note that this high volume is counterbalanced by a high degree of spatio-temporal correlations: for instance, pixels in a video sequence do not evolve independently. This feature can be exploited to substantially reduce the dimensionality of the problem by embedding the raw data in low dimensional manifolds. Since the projections to/from these manifolds can be modeled as memoryless (possibly, time varying) nonlinearities, this approach leads, locally, to the Hammerstein/Wiener system identification problem illustrated in Fig. 3. Motivated by polynomial kernel embeddings, in order to solve the problem, the given temporal data $\{y_t\}$, was embedded, via a nonlinear projection $\xi_t = \Pi(y_t)$ in manifolds where its evolution could be (locally) explained by a linear model of the form: $\xi_k = \sum_{i=1}^{n_a} a_i\xi_{k-i} + \eta_k$, where η_k accounted for approximation error. Next, to each embedded time series, we associated a Hankel matrix of the form:

$$H_\xi \doteq \begin{bmatrix} \xi_0 & \xi_1 \cdots & \xi_m \\ \xi_1 & \xi_2 \cdots & \xi_{m+1} \\ \vdots & \vdots \ddots & \vdots \\ \xi_n & \xi_{n+1} \cdots & \xi_{m+n} \end{bmatrix}$$

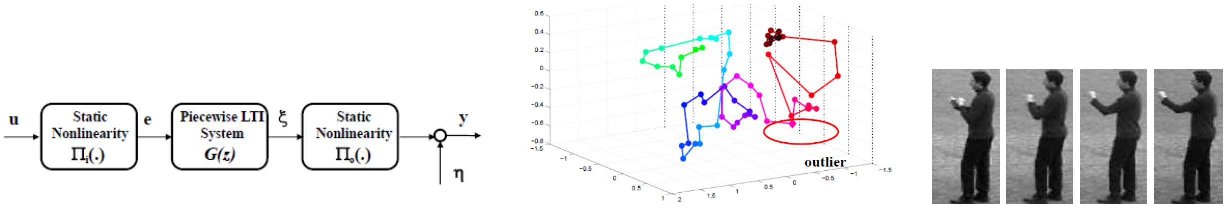


Figure 3: Left: Manifold embedding as a nonlinear identification: Here $\Pi_i(.)$ and $\Pi_o(.)$ are memoryless nonlinearities, $(u \in R^{n_u}, y \in R^{n_y})$ and $(e \in R^{n_e}, \xi \in R^{n_\xi})$, with $n_u \gg n_e$ and $n_y \gg n_\xi$ represent the (high dimensional) raw data and its projection on the low dimensional manifold, respectively, and the piecewise linear dynamics $G(.)$ governs the evolution of data on the manifold. Center: 3 dimensional manifold obtained applying our approach to the KTH boxing sequence shown in the right. Note that the outlier has minimal influence on the manifold structure.

Since the the vector $w \doteq \{a_1, \ldots, a_{n_a}\}$ satisfies $H_\xi w = 0$, it follows that the dynamic data is completely characterized by the null space of its associated Hankel matrix. Inspired by maximum margin classification algorithms and the result above, we developed a maximum margin dynamics-based classification algorithm that worked with the null spaces of Hankel matrices [15]. Given two sets of training data $\{y_t^+\}$ and $\{y_t^-\}$ corresponding to nominal and anomalous scenarios, we jointly sought for embeddings ξ_t^+, ξ_t^- and a vector w such that minimizes γ subject to $\|H_{\xi^+} w\|_2^2 \leq \gamma$ and $\|H_{\xi^-} w\|_2^2 > 1 + \gamma$. Intuitively, we sought a vector w such that (i) it approximately lied in the null space of the Hankel matrices of all the positive examples dynamic sequences, and (ii) it maximized the margin between the residue $\|H_{\xi^+} w\|_2^2$ for the nominal and anomalous sequences. Defining the Kernel (or Gram) matrix by its submatrices

$$K_{i,j} = \begin{bmatrix} \xi_j \xi_j & \xi_j \xi_{j+1} & \cdots & \xi_j \xi_{j+c} \\ \xi_{j+1} \xi_j & \xi_{j+1} \xi_{j+1} & \cdots & \xi_{j+1} \xi_{j+c} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{j+c} \xi_j & \xi_{j+c} \xi_{j+1} & \cdots & \xi_{j+c} \xi_{j+c} \end{bmatrix}$$

and noting that $G_i = H_i^T H_i = \sum_{j=1}^{T-c+1} K_{i,j}$, allowed to reduce the problem outlined above to (see [15] for details):

$$\begin{array}{l} \min_{K,w,\gamma} \frac{1}{2}||w||_2^2 + C\gamma - \lambda \text{Trace}(K) \\ \text{subject to: } w^T G_i w \leq \gamma \quad, \forall G_i \in G_+ \\ w^T G_i w + \gamma \geq 1 \quad, \forall G_i \in G_- \\ G_i = \sum_{j=1}^{T-c+1} K_{i,j} \\ K \succeq 0, \ \gamma \geq 0 \\ (1-\epsilon)||y_i - y_j||^2 \leq k_{ii} + k_{jj} - 2k_{ij} \leq (1+\epsilon)||y_i - y_j||^2 \end{array} \tag{10}$$

where the last constraint approximately enforced preservation of the local spatial geometry and where the additional term $-\lambda \text{Trace}(K)$ in the objective sought to favor lower dimensional embeddings. The semi-algebraic problem was solved by using recent results in sparse polynomial optimization, that exploited its inherent sparsity to substantially reduce the computational burden [15].

2.1.4 Model (In)Validation: Consider now the problem of establishing whether a noisy input/ouput sequence could have been generated by a given model of the form (2), possibly subject to model uncertainty. Classically, model (in)validation has been used as an intermediate step following identification and prior to use the identified models for control synthesis. Interestingly, as we established in the course of this research, the same ideas can be used in the context of information extraction to identify contextually abnormal sequences (see section 2.2.3). Formally, the problem of interest can be stated as establishing whether a noisy input/ouput sequence could have been generated by a given model of the form:

$$\begin{array}{rcl} y_t & = & \sum_{i=1}^{n_a} a_i(\sigma_t) y_{t-i} + \sum_{i=1}^{n_c} c_i(\sigma_t) u_{t-i} \\ \tilde{y}_t & = & y_t + \eta_t, \ \sigma_t \in \{1, \ldots, s\}, \ \|\eta_t\|_\infty \leq \epsilon \end{array} \tag{11}$$

where $\tilde{y}_t$ denotes the measured output corrupted by the noise η_t. As in the identification case, this problem is known to be generically NP-hard, due to the presence of noise and the fact that the mode variable σ_t is not directly measurable. Cases where σ_t takes only a small number of discrete values (for instance a system switching between two known modes), can be handled by simply considering all possible switching sequences. Clearly, due to its combinatorial nature, this approach becomes infeasible for cases involving relatively small number of subsystems. On the other hand, this combinatorial complexity can be avoided by appealing to semi-algebraic geometry tools. To this effect, begin by noting that, (11) holds if and only if:

$$p_r(\eta_{t:t-n_c}) \doteq \prod_{i=1}^{s} [g_{t,i}(\eta_{t:t-n_c})]^2 = 0 \tag{12}$$

where:

$$\begin{array}{l} g_{t,i}(\eta_{t:t-n_c}) \doteq a_1(i)(\tilde{y}_{t-1} - \eta_{t-1}) + \ldots + a_{n_a}(i)(\tilde{y}_{t-n_a} - \eta_{t-n_a}) - (\tilde{y}_t - \eta_t) + c_1(i)u_{t-1} \\ + \ldots + c_{n_c}(i)u_{t-n_c} \end{array} \tag{13}$$

Similarly, the norm constraint on the noise sequence η_t is equivalent to the polynomial constraint $h_t(\eta_t) \doteq \epsilon^2 - \eta_t^2 \geq 0$. Hence, there exists noise and switching sequences such that (12) holds if and only if the semi-algebraic set

$$\mathcal{T}(\eta) \quad \doteq \quad \{\eta \mid f_t(\eta_t) \geq 0 \; \forall \; t \in [t_o, T] \text{ and } p_t(\eta_{t:t-n_a}) = 0 \; \forall t \in [n_a, T]\}\,. \tag{14}$$

is not empty. Thus, an (in)validation certificate can be obtained by considering the following optimization problem:

$$\begin{array}{l} o^* = \min_\eta \sum_{t=n_a}^{T} p_t(\eta_{t:t-n_a}) \\ \text{s.t. } f_t(\eta_t) \geq 0 \;\; \forall t \in [0, T] \end{array} \tag{15}$$

Note that $o^* > 0 \iff \mathcal{T}'(\boldsymbol{\eta}) = \emptyset$. While computing o^* requires solving a computationally challenging polynomial optimization problem, a convergent sequence of lower bounds can be obtained using recent results from semi-algebraic optimization, leading to a sequence of problems of the form:

$$\begin{array}{ll} d_N^* = & \min_{\mathbf{m}} \sum_{t=n_a}^{T} l_t(\mathbf{m}_{t-n_a:t}) \\ & \text{s.t.} \\ & \mathbf{M}_N(\mathbf{m}_{t-n_a:t}) \succeq 0 \; \forall t \in [n_a, T] \\ & \mathbf{L}_N(f_t \mathbf{m}_{t-n_a:t}) \succeq 0 \;\; \forall t \in [n_a+1, T] \end{array} \tag{16}$$

where l_t is a linear functional of the optimization variables $\mathbf{m}$ and where $\mathbf{M}_N$ and $\mathbf{L}_N$ are matrices affine in these variables. Hence, these problems can be efficiently solved by using commonly available semi-definite optimization solvers. It is worth emphasizing that this reformulation allows for exploiting the inherently sparse structure of the problem, resulting in substantial computational complexity reduction [11,28].

2.1.5 Robust estimation under ℓ^∞ bounded disturbances. Traditional noise models often do not capture key features of the problems of interest here. As a simple example, noise in images should be bounded. While in principle this feature can be captured using truncated distributions, the resulting problems are computationally hard. To circumvent this difficulty we developed a new framework for robust estimation in the presence of unknown-but-bounded noise. Using a concept similar to superstability led to robust filters that can be synthesized by simply solving a linear programming problem [1]. A salient feature of this framework is that it explicitly allows for trading off filter complexity against worst-case estimation error. We have also extended this framework to the more challenging case where the mode variable is not accessible to the filter and shown that the resulting problem can be recast into a (polynomial) sparsification form and

solved using results from semi-algebraic geometry [13]. In addition, we explored a new information-based complexity framework that combines the properties of traditional worst case and probabilistic estimation approaches, leading to a substantial reduction in the conservatism of the former, while retaining its ability to provide bounds on worst case errors [8,9,27]. Finally, we have developed computationally tractable algorithms for synthesizing optimal filters subject to sparsity constraint on the information flow, and for optimal sensor selection. The main result of [19] showed that, surprisingly, the first problem is convex, while in [16] we showed that while the second problem in non-convex, tractable convex relaxations with optimality guarantees can be obtained using tools from semi-algebraic geometry.

2.2 Application: Detecting Contextually Abnormal Events:

We applied the theoretical framework described in section 2.1 to a problem at the core of flexible autonomy: detecting contextually abnormal activities. Solving this problem in realistic, potentially adversarial environments required the ability to (i) perform persistent tracking, (ii) detect significant events, and (iii) recognize activities from noisy, fragmented data records. As briefly outlined below, the framework developed in this research indeed leads to robust, computationally tractable solutions to these problems.

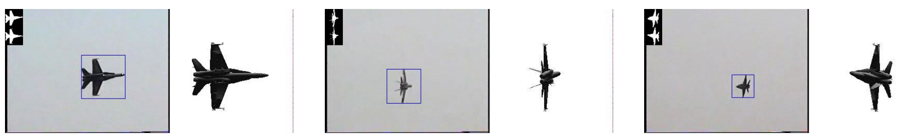


Figure 4: Using a Hammerstein/Wiener system to achieve sustained tracking in the presence of appearance changes. The left and right portions of each frame show the actual and predicted target appearance, with their correlation displayed in the top left corner.

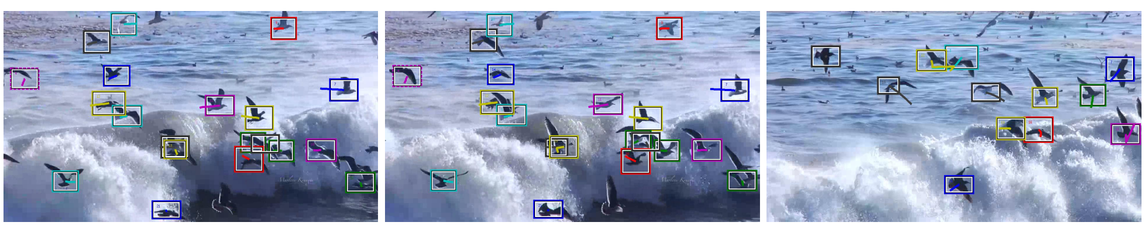


Figure 5: Using dynamics to track targets with similar appearance

2.2.1 Tracking via Robust Nonlinear Operator Embeddings. The ability to persistently track and disambiguate is a key enabler for flexible autonomy. However, this process is far from trivial in urban environments due to occlusion and target appearance changes, compounded by the (potential) existence of multiple targets with similar appearance. In the context of this research, we overcame these barriers by using our identification framework to map the data (in this case image features) to points on low dimensional manifolds where dynamics are locally linear time invariant, effectively decoupling appearance (embedded in the manifold structure) from intrinsic dynamics. The resulting models were used to reconstruct missing data, predict future target positions and disambiguate targets. The potential of this approach is illustrated in Figure 4, where it was used to achieve sustained tracking in the presence of extreme appearance changes, due to a target U-turn. In addition, exploiting the dynamical information allowed for sustained tracking under substantial occlusion [6], and for disambiguating multiple targets with similar appearances, such as those shown in Fig. 5 [18].

2.2.2 Dynamic data segmentation and event detection. These problems can be embedded in the proposed identification framework by simply noting that events correspond to mode changes in the underlying dynamical system and thus can be detected by monitoring changes in invariants associated with individual models. The simplest such invariant is model order, since, intuitively, models associated with homogeneous data, e.g. a single activity, have far lower complexity than those jointly explaining multiple datasets. Boundaries are thus characterized by an increase in model complexity, and can be detected by performing a sequence of SVDs of empirical Hankel matrices. An example of the potential of this approach to detect activity changes from real video feeds is shown in Fig. 6.

Figure 6: . Fast event detection. The jump in the rank of the Hankel matrix indicates a change in dynamics as the suspect of a 2010 bombing attack in New York City stops to remove his sweater.

2.2.3 Activity Recognition and Anomaly Detection: The vision driving the application domains considered in this proposal is that of autonomous systems endowed with the capability to recognize anomalous behavior. We propose to embed this problem into our identification/model (in)validation framework as follows. The starting point is to consider activities as second-order stationary stochastic processes. Thus, each activity can be considered as the output of a time-invariant dynamical system. Further, by projecting the raw data into suitable manifolds allows for decoupling the effect of "nuisance" factors (such as view-point or appearance changes) from the intrinsic dynamics of the activity under consideration. Then, given a sequence of frames from a single unknown activity, recognition can be accomplished by interrogating a database of known activity models to establish whether it contains an element (and an associated uncertainty description) compatible with the observed data. A difficulty here is that a single activity can consist of the concatenation of several sub-activities of various lengths. For instance a "normal" activity could consist of walking for two minutes, standing for one, and then resuming walking. However, this is precisely the situation addressed by the proposed switched model (in)validation framework described in Section 2.1.4. Advantages of this (in)validation based approach over existing ones include the ability to fully exploit dynamic information, handle data streams that do not overlap in time and directly eliminate the effect of nuisance factors. These advantages are illustrated in Fig. 7 using a simple example involving two known activities.

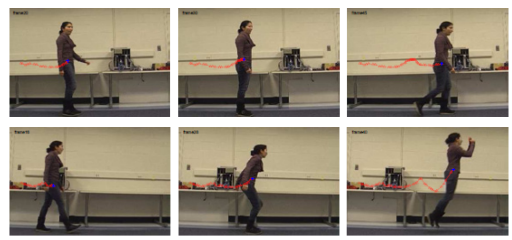


Figure 7: Anomalous behavior detection as a switched (in)validation problem. The activity database consists of models of two activities, "walk" and "wait". The top sequence (walk–wait–walk) is not (in)validated since both activities are in the database. The bottom sequence (walk-jump) is flagged as abnormal since it cannot be generated by switching amongst models in the database.

2.2.4 Finding Causal Interactions in Video Data. In many scenarios, seemingly benign individual actions can indeed aggregate to potential threats. An example of these situations are flash mobs. Thus, as part of this research we applied our identification framework to the problem of detecting causally interacting individuals. The main idea, illustrated in Fig. 8 is to recast the problem into a sparse dynamical graphical model identification form. In this context, each node corresponds to the observed motion of a given target, and each link indicates the presence of a causal correlation. As we showed in [17], this approach led to a block-sparsification problem that can be efficiently solved using a modified Group-Lasso type approach, capable of handling missing data and outliers (due for instance to occlusion and mis-identified correspondences).

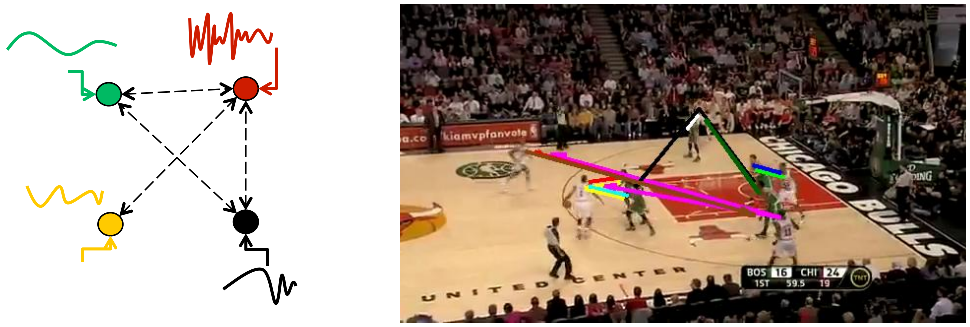


Figure 8: Finding causal interactions as a graph identification problem. Left: Representation of this sequence as a graph, where each node represents the time series associated with the position of each player and the links are vector regressive models. Causal interactions exist when one of the time series can be explained as a combination of past values of the others. Right: Application of these ideas to the problem of finding causally interacting players in a basketball game.

Moreover, this approach also identified time instants where the interactions between agents changed, thus providing event detection capabilities. Efficient computational methods were developed by combining this idea with the parsimonious model identification framework developed in [4, 20,29].

2.3 Personnel Supported During the Duration of the Grant

Mario Sznaier	Dennis Picard Trustee Professor of ECE, Northeastern University (PI)
Octavia Camps	Professor of ECE, Northeastern University (Co-PI)
Burak Yilmaz	Ph.D. student, graduated, 2014
Yongfang Cheng,	PhD. student, expected graduation date, May 2015
Jose Lopez	Ph.D. student, expected graduation date, May 2015
Yin Wang	Ph.D. student, expected graduation date, May 2015

2.4 Honors and Awards Received

- Mario Sznaier was awarded a Distinguished Member Award by the IEEE Control Systems Society (less than 100 people have received this award since the society was created)
- Mario Sznaier delivered plenary talks at the 2012 IFAC Symposium on Robust Control Design, 2012 IFAC Symposium on System Identification, the 2012 Mediterranean Control Conference, the 2013 RPIC, the 2015 Symposium on Data Science and Systems Complexity, and the 2015 Geometric and Numerical Foundations of Movement; and a Semi-Plenary lecture at the 2012 IEEE Conference on Decision and Control.
- Octavia Camps was a keynote speaker at the 2014 IEEE International Conference on Distributed Smart Cameras.

2.5 Transitions

The theoretical framework developed under this grant was used to develop a “contra-flow” detector to support TSA agents by alerting them to potential attempts to breach secure areas. This technology was tested at the Cleveland Hopkins International Airport for over a year, were it screened on average 50,000 passengers per week. This technology was showcased to the Hon. Janet Napolitano (US secretary of Homeland Security, November 2012), Mr. John S. Pistole (TSA administrator and former FBI Deputy Director, June 2013) and the Hon. Theresa May (U.K. Home Secretary, Sept. 2014). It was also covered in a N.Y. Times article that appeared on May 8, 2015.

Figure 9: Left: security breach detection technology being demonstrated to the Hon. J. Napolitano, Homeland Security Secretary. Right: technology demo for Mr. J. Pistole, TSA Administrator and former FBI Deputy Director.

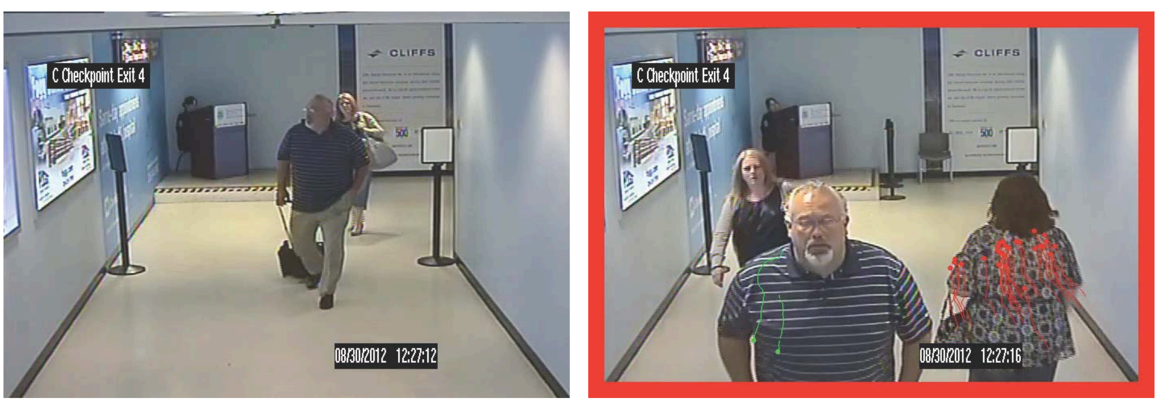


Figure 10: Security breach detection technology deployed at the Cleveland Hopkins international airport.

2.6 Disclaimer

The views and conclusions contained in this report are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFOSR or the U.S. Government.

2.7 Publications

1. F. Blanchini and M. Sznaier, A Convex Optimization Approach to Synthesizing Bounded Complexity ℓ^∞ Filters, IEEE Trans. Aut. Contr, pp. 216 221, 2012.

2. N. Ozay, M. Sznaier, C. Lagoa, and O. Camps, A Sparsification Approach to Set Membership Identification of Switched Affine Systems, IEEE Trans. Aut. Contr, pp. 634–648, 2012.

3. B. Li, O. Camps, and M. Sznaier, Cross-view Activity Recognition using Hankelets, the 2012 IEEE Conf. Comp. Vision and Pattern Recognition, 2012 (acceptance rate 24%), pp. 1362–1369.

4. M. Ayazoglu, M. Sznaier, and O. Camps, Fast Algorithms for Structured Robust Principal Component Analysis, the 2012 IEEE Conf. Comp. Vision and Pattern Recognition, 2012 (acceptance rate 24%), pp. 1704–1711.

5. C. Feng, C. Lagoa, and M. Sznaier, A Convex Approach to Generalized Fixed Order Interpolation, Proc. 2012 American Control Conference, 2012, pp. 5174–5171

6. F. Xiong, O. Camps and M. Sznaier, Dynamic Context for Tracking Behind Occlusions, Proc. 2012 ECCV, pp. 580 - 593 (acceptance rate 25%).

7. M. Sznaier, Compressive Information Extraction: A Dynamical Systems Approach," Proc. 2012 Symposium on Systems Identification (SYSID 2012), Plenary Paper, pp. 1559–1568, 2012.

8. Dabbene, F., Sznaier, M., andTempo, R., A Probabilistic Approach to Optimal Estimation. Part I: Problem Formulation and Methodology, Proc. 2012 IEEE Conf. Dec. and Control, pp. 190–195.

9. Dabbene, F., Sznaier, M., andTempo, R., A Probabilistic Approach to Optimal Estimation. Part II: Algorithms and Applications,Proc. 2012 IEEE Conf. Dec. and Control, pp. 196–201.

10. Ayazoglu. M., and Sznaier, M., An Algorithm for Fast Constrained Nuclear Norm Minimization and Applications to Systems Identification, Proc. 2012 IEEE Conf. Dec. and Control, pp. 3469–3475.

11. Cheng, Y., Wang, Y., Sznaier, M., Ozay, N., and Lagoa, C., A Convex Optimization Approach to Model (In)validation of Switched ARX Systems with Unknown Switches, Proc. 2012 IEEE Conf. Dec. and Control, pp. 6284–6290.

12. Bekiroglu, K., Sznaier, M., Lagoa, C. and Shafai, B., Vision Based Control of an Autonomous Blimp with Actuator Saturation Using Pulse Width Modulation, Proc. 2013 IEEE Conf. Control Applications, pp. 1036–1041.

13. Cheng, Y., Wang, Y. and Sznaier, M. Worst Case Optimal Estimators for Switched Linear Systems, Proc. 52 IEEE Conf. Dec. Control, Dec. 2013, pp. 4036–4041.

14. Yilmaz, B., Lagoa, C. and Sznaier, M. An Efficient Atomic Norm Minimization Approach to identification of Low Order Models, Proc. 52 IEEE Conf. Dec. Control, Dec. 2013, pp. 5834–5839.

15. Xiong, F., Cheng, Y., Camps, O., Sznaier, M. and Lagoa, C., Hankel Based Maximum Margin Classifiers: A Connection Between Machine Learning and Wiener Systems Identification, Proc. 52 IEEE Conf. Dec. Control, Dec. 2013, pp. 6005–6010

16. Wang, Y., Sznaier,M. and Dabbene, F., A Convex Optimization Approach to Worst Case Optimal Sensor Selection, Proc. 52 IEEE Conf. Dec. Control, Dec. 2013, pp. 6353–6358

17. Ayazoglu, M., Yilmaz, B., Sznaier, M. and Camps, O. Finding Causal Interactions in Video Sequences, IEEE Int. Conf. on Computer Vision, Sydney, Australia, Dec. 2013.

18. Dicle, C., Camps, O., and Sznaier, M. The Way They Move: Tracking Multiple Targets with Similar Appearance, IEEE Int. Conf. on Computer Vision, Sydney, Australia, Dec. 2013. (Oral, acceptance rate 2.52%)

19. Lopez, J., Wang, Y., and Sznaier, M. Sparse $\mathcal{H}_2$ Optimal Filter Design via Convex Optimization, 2014 American Control Conference, pages 1108–1113

20. K Bekiroglu, B Yilmaz, C. Lagoa, and M. Sznaier. Parsimonious model identification via atomic norm minimization. In 2014 European Control Conference, pages 2392–2397, 2014.

21. F. Xiong, M. Gou, O. Camps and M. Sznaier, Person Re-Identification Using Kernel-Based Metric Learning Methods, 2014 European Comp. Vision Conf, Zurich, Switzerland, pp. 1–16, Sept. 2014 (28% acceptance rate)

22. W. Lucia, M. Sznaier and G. Franze, An obstacle avoidance and motion planning Command Governor based scheme: The Qball-X4 quadrotor case study, Proc 53rd IEEE CDC, Page(s): 6135–6140, 2014

23. Y. Wang, J. Lopez and M. Sznaier, Sparse static output feedback controller design via convex optimization, Proc 53rd IEEE CDC, Page(s): 376–381, 2014.

24. Y. Cheng, Y. Wang, M. Sznaier, A convex optimization approach to semi-supervised identification of switched ARX systems, Proc 53rd IEEE CDC, Page(s): 2573–2578, 2014

25. M. Sznaier , O. Camps, N. Ozay and C. Lagoa,Surviving the upcoming data deluge: A systems and control perspective, (invited tutorial paper) Proc 53rd IEEE CDC, Page(s): 1488–1498, 2014.

26. L. Lo Presti, M. La Cascia, S. Sclaroff and O. Camps, Gesture Modeling by Hanklet-based Hidden Markov Model. Proc. Asian Conference on Computer Vision (ACCV), Nov. 2014 (28% acceptance rate)

27. F. Dabbene, F. M. Sznaier and R. Tempo, Probabilistic Optimal Estimation with Uniformly Distributed Noise. IEEE Trans. Autom. Contr..59(8), 21132127, 2014.

28. N. Ozay, N. M. Sznaier, and C. Lagoa,Convex Certificates for Model (In)validation of Switched Affine Systems with Unknown Switches, IEEE Trans. Autom. Contr..59(11), 2921–2932, 2014

29. M. Sznaier, M. Ayazoglu,and T. Inanc, Fast Structured Nuclear Norm Minimization with Applications to Set Membership Systems Identification, IEEE Trans. Autom. Contr..59(10), 2837–2842, 2014.

30. Y. Wang, C. Dicle, M. Sznaier and O. Camps,Self Scaled Regularized Robust Regression, IEEE CVPR, Page(s) 3261–3269, 2015 (acceptance rate 28%)

31. Y. Cheng, J. Lopez, O. Camps and M. Sznaier, ”A convex optimization approach to robust fundamental matrix estimation, IEEE CVPR, Page(s) 2170–2178, 2015 (acceptance rate 28%)

32. N. Ozay, C. Lagoa and M/ Sznaier, Set Membership Identification of Switched Linear Systems with Known Number of Subsystems, Automatica, page(s)180–191, 2015